

Application/Control Number: 10/662,550  
Art Unit: 2628

Docket No.: 2000-0042-CON

RECEIVED  
CENTRAL FAX CENTER

AUG 18 2006

REMARKS

Reconsideration and allowance in view of the foregoing amendment and the following remarks are respectfully requested. Claims 22, 23 and 30 are amended without prejudice or disclaimer. Applicant reserves the right to pursue broader claims in a continuation application.

Rejection of Claims 22-25, 27, 29-32 and 34 Under 35 U.S.C. §103(a)

The Office Action rejects claims 22-25, 27, 29-32 and 34 under 35 U.S.C. §103(a) as being unpatentable over Ezzat et al. ("Visual Speech Synthesis by Morphing Speech Analysis") ("Ezzat et al.") in view of Jiang et al. ("Visual Speech Analysis with Application to Mandarin Speech Training") ("Jiang et al.") in view of Bregler et al. We note that Applicant does not acquiesce to any broadening of the teachings of Bregler et al. as being admitted prior art. For example, the specification notes on page 2 that Bregler et al. utilize triphone segments as the a priori units of video, thus causing an lack of natural flow. Thus, Applicant does not admit that the breadth of Bregler et al. extends beyond the teachings of using triphone segments as the a priori units of video.

Applicant first traverses the combination of Ezzat et al., Jiang et al. and Bregler et al. For example, the office action states that it would be obvious to combine Ezzat et al. with Jiang et al., but give no analysis of why other than to state that Jiang et al. teach an advantage of obtaining feature vectors to help children improve speech pronunciations. The Examiner has the burden of establishing a *prima facie* case of obviousness and Applicant submits that such a case has not been made. The reason one of skill in the art would not be motivated to combine these references is that Ezzat et al.'s disclosure relates to the process of providing speech synthesis by morphing visemes. As stated in the Abstract, visemes are a small set of images spanning a large range of mouth shapes. The viseme database is created by recording a human subject in a manner specifically designed to elicit one instantiation of each viseme. One of skill in the art

BEST AVAILABLE COPY

Application/Control Number: 10/662,550  
Art Unit: 2628

Docket No.: 2000-0042-CON

will understand the technical process in generating a viseme database and the process of morphing those visemes to create a smooth animation.

Jiang et al. on the other hand, discloses a speech analysis system for teaching mandarin Chinese to oral deaf children. The portion of the abstract cited in the Office Action, "at each frame, region of interest is identified and key information is extracted," clearly is in the context of a camera that video tapes a child speaking in Chinese such that the child can be video taped and the language analyzed. Later, a talking head (which is a representation of the child) is presented to the child to help with correcting errors. The talking head recreated here is a representation of the child and made from the child's own images. There is no discussion of visemes or suggestion that the cloned talking head is presented using a viseme process. Therefore, Applicant submits that one of skill in the art would recognize the technical differences and would not be motivated to use Jiang et al.'s feature vectors because there is no hint that they can apply to visemes.

Next, Applicant submits that claims 22, 23 and 30 as amended recite limitations that overcome the prior art even if combined. For example, claim 22 recites a method for the synthesis of photo-realistic animation of an object. The method includes the step of selecting candidate image samples utilizing the target feature vector to generate a photo-realistic animation of the object, wherein generating the photo-realistic animation of the object occurs uses an audio/video unit selection process in which a longest possible candidate image sample is selected. There is no teaching within the prior art of record regarding a selection process in which the longest possible candidate image sample is selected. In Ezzat et al., for example, they teach that the visual corpus is digitized at 30 fps. Ezzat has the need for dealing with many intermediate frames that lie between the chosen viseme images. Accordingly, Applicant

BEST AVAILABLE COPY

Application/Control Number: 10/662,550  
Art Unit: 2628

Docket No.: 2000-0042-CON

submits that claim 22, claim 23 and its dependent claims and claim 30 and its dependent claims are patentable over the cited references for the several reasons set forth above.

**Rejection of Claims 28 and 35 Under 35 U.S.C. §103(a)**

The Office Action rejects claims 28 and 35 under 35 U.S.C. §103(a) as being unpatentable over Ezzat et al. in view of Jiang et al. in further view of Bregler et al. in further view of Brand ("Voice Puppetry") ("Brand"). Applicant traverses this rejection and submits that Brand fails to teach the limitations of claims 28 and 35.

The Office Action equates page 25, column 1's discussion in Brand regarding the Viterbi sequence with the step in claim 28 of selecting for each frame a number of candidate image samples from the first database based on the target feature vector. The parent claim 22 requires the target feature vector to have a visual and a non-visual aspect to it. This feature is not taught in Brand. Further, Brand fails to teach the step of calculating a concatenation cost from a combination of visual features from the second database and object characteristics from the third database. Therefore, even if combined, Brand fails to teach each limitation of claim 28 as well as claim 35.

Furthermore, Brand teaches away from its combination with the other references. In fact, Brand expressly distances itself from the Bregler et al. reference. Bregler et al. is listed as citation [7] in Bregler et al. On page 22, top of col. 1, Brand states that nearly all lip-syncing systems are based on an intermediate phonemic representation whether obtained by hand, text or via speech recognition (citing Bregler et al. and others). Brand cites Bregler et al. again in that paragraph noting that

"[7] works by re-ordering existing video frames rather than by generating animations, it deserves mention because it partially models vocal (but not facial) co-articulation with triphones – phonemes plus one 1 unit of left and right context. The quality of a video rewrite is determined by the amount of video that is available to be provide [sic] triphone examples and how successfully it is analyzed; smoothing is necessary because triphones

BEST AVAILABLE COPY

Application/Control Number: 10/662,550  
Art Unit: 2628

Docket No.: 2000-0042-CON

don't fully constrain the solution and no video will provide an adequate stock of triphones.

...  
None of these methods address the actual dynamics of the face."

Accordingly, it is clear that Brand makes special mention of the Bregler et al. article in the background section to highlight it as a reference which fails to "address the actual dynamics of the face." In this respect, the Bregler et al. article is cast as teaching undesirable techniques to performing animation. In this regard, Applicant respectfully submits that there cannot be any motivation or suggestion to combine these two references where there are express teachings away from such combination.

Similarly, Brand discloses in the same part of page 22 that references that use "visemic tokens" are also deficient in that they fail to address the actual dynamics of the face.

Accordingly, one of skill in the art would recognize that Brand teaches away from using the viseme approach taught by Ezzat et al. As another example of how Brand distances his teachings from a viseme-based approach, he states "Considerable information can be lost when discretizing to phonemic or visemic representations." Col. 1, page 22. Accordingly, Brand also cannot be combined with Ezzat et al. because if his express teachings away from such a combination.

Based on the foregoing reasons, Applicant submits that claims 28 and 35 are patentable and in condition for allowance.

BEST AVAILABLE COPY

Application/Control Number: 10/662,550  
Art Unit: 2628

Docket No.: 2000-0042-CON

RECEIVED  
CENTRAL FAX CENTER

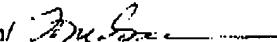
CONCLUSION

AUG 18 2006

Having addressed all rejections and objections, Applicant respectfully submits that the subject application is in condition for allowance and a Notice to that effect is earnestly solicited. If necessary, the Commissioner for Patents is authorized to charge or credit the Isaacson, Irving, Stelacone & Prass, LLC, Account No. 502960 for any deficiency or overpayment.

Respectfully submitted,

Date: August 18, 2006

By: Thomas M. Isaacson 

Correspondence Address:  
Thomas A. Restaino  
Reg. No. 33,444  
AT&T Corp.  
Room 2A-207  
One AT&T Way  
Bedminster, NJ 07921

Thomas M. Isaacson

Attorney for Applicant  
Reg. No. 44,166  
Phone: 410-286-9405  
Fax No.: 410-510-1433

BEST AVAILABLE COPY